The aim of this study was to assess the reliability, validity and feasibility of selected physical performance tests, to compile a new fitness test battery based on these results and to obtain standard values for young men. 79 men (20.3 ± 1.1 y) performed the tests for the reliability part, while 60 men (20.3 ± 1.1 y) completed the tests for the validity part of the study. Feasibility was confirmed by 25 sport experts who conducted the test battery among 1704 men (19.5 ± 1.0 y). For standard values, the data of 12,862 men (19.9 ± 1.0 y) were collected. Based on the reliability and validity data, the following 5 tests were selected for the fitness-test battery: 1) progressive endurance run, 2) seated 2-kg-shot put, 3) standing long jump, 4) trunk muscle strength test and 5) 1-leg standing test. The reliability and validity of the selected performance tests were sufficient to very good (r = 0.50–0.90 and r = 0.64–0.91, respectively). The suggested fitness-test battery can be applied among large groups.

Assembling and Verification of a Fitness Test Battery for the Recruitment of the Swiss Army and Nation-wide Use

**Abstract**

The aim of this study was to assess the reliability, validity and feasibility of selected physical performance tests, to compile a new fitness test battery based on these results and to obtain standard values for young men. 79 men (20.3 ± 1.1 y) performed the tests for the reliability part, while 60 men (20.3 ± 1.1 y) completed the tests for the validity part of the study. Feasibility was confirmed by 25 sport experts who conducted the test battery among 1704 men (19.5 ± 1.0 y). For standard values, the data of 12,862 men (19.9 ± 1.0 y) were collected. Based on the reliability and validity data, the following 5 tests were selected for the fitness-test battery: 1) progressive endurance run, 2) seated 2-kg-shot put, 3) standing long jump, 4) trunk muscle strength test and 5) 1-leg standing test. The reliability and validity of the selected performance tests were sufficient to very good (r = 0.50–0.90 and r = 0.64–0.91, respectively). The suggested fitness-test battery can be applied among large groups.

**Introduction**

In terms of public health, the monitoring of the population’s fitness level is important because it is positively related to health outcomes. Since 1874 a physical performance test has been part of the recruitment process in the compulsory Swiss Army. Every 19-year-old male Swiss citizen has to participate. The former performance tests have never been evaluated scientifically. The need for a feasible, valid and standardised physical fitness test battery for population monitoring, education and occupational medicine has been growing lately in Switzerland. The previously used and other known physical fitness test batteries are either too time-consuming, or they contain tests with limited reliability or validity. Therefore, a Swiss physical fitness test battery (SPFTB) for nation-wide use in young men was developed. SPFTB should a) permit an evaluation of large groups of young men in a limited time period, b) be feasible with a minimum of material, c) comply with scientific criteria of validity and reliability, and d) contain relevant performance- and health-related components of physical fitness.

In 1951, Cureton (1951) defined physical fitness as the degree of balance, flexibility, agility (speed), strength, power and endurance. Several definitions of physical fitness have since been published. Miller et al. (1991) defined physical fitness in more general terms as the level of ability to perform sustained physical work characterized by an effective integration of cardiorespiratory endurance, strength, flexibility, coordination and body composition. Physical fitness is a multidimensional construct and therefore cannot be assessed by a single test. Thus a battery of different tests is needed. According to our definition of physical fitness, the test battery should measure health-related factors and aspects of performance. A complete fitness-test battery should therefore assess a) cardiorespiratory endurance, b) muscle strength and endurance, and c) agility and balance.

A reliable and valid measure of cardiorespiratory endurance is the maximal oxygen consumption (VO2max; Safrit et al., 1988), although it is not immune to inaccuracy (Shepard, 1984). In larger population groups, direct measurement of VO2 is not feasible since expensive equipment is required. Therefore, a variety of less complex tests to measure cardiorespiratory endurance was developed. The frequently used 12-min run test (12-MRT), also referred to as Cooper test, is strongly related to the criterion measure of VO2max in adults (r = 0.84–0.92; Cooper, 1968; Grant et al., 1995; McCutcheon et al., 1990). The 12-MRT is appropriate for individuals with a sufficient fitness level and requires considerable motivation and experience for self-pacing. The multistage 20-m shuttle run test (MST) is more appropriate for individuals without experience for self-pacing. The MST paces the participants by an acoustic signal and has shown to be an accurate method to estimate VO2max in adults (r = 0.79–0.90; Cooper, 1968; Grant et al., 1995; Leger et al., 1989; McNaughton et al., 1998; Ramsbottom et al., 1988).
However, the frequent stopping and starting may limit the applicability for individuals with less developed motor skills, especially at higher speed levels. A paced progressive endurance run on a track could be a more appropriate method to measure cardiorespiratory endurance among heterogeneous population groups.

To estimate the muscle power in the upper and lower extremities, peak power during bench press throw, squat jump (SJ) and counter movement jump (CMJ) are evaluated. In larger population groups, this direct measurement of muscle power is not feasible. Therefore, less complex tests were developed to measure muscle power. The simple shot put is a feasible and valid test to measure the power of upper extremities in larger groups. The seated 4.5-kg-shot put, conducted by members of a weight training class, correlated positively (r = 0.75) with their power during bench press throw of 60% of their 1-repetition maximum (Mayhew et al., 1991). The distance of the seated 0.4 kg chest pass, conducted by women of a netball team, correlated significantly with peak power during bench press throw of 10 kg (r = 0.80; Cronin and Owen, 2004). Based on these results, we suggest that a weight of 2 kg could be appropriate for seated shot put in a heterogeneous population of young men.

To measure the power of lower extremities, high or long jumps are feasible tests. The vertical jump-and-reach score is a good predictor for the power in the leg extensor muscles (r = 0.93 resp. r = 0.91; Sayers et al., 1999). The standing long jump (SLJ) is widely used because of the good feasibility and test-retest reliability (r = 0.89–0.95; Markovic et al., 2004; Tsigilis et al., 2002). The correlation between SLJ and the principal component of explosive power is good (r = 0.76; Markovic et al., 2004). Concerning the assessment of maximal running speed, a pendulum sprint is a well-known method, which can be applied in a gym hall. However, its validity has yet to be demonstrated.

Sit-ups are often used to measure the muscular strength and endurance of the abdominal muscle groups. Some studies showed limited reliability of dynamic or isometric sit-up tests (r < 0.50; Sparling et al., 1997; Suni et al., 1996), while others indicated satisfactory reliability (r = 0.72–0.84; DiNucci et al., 1990; Erbaugh, 1990; Tsigilis et al., 2002). Available data suggest that sit-ups yield limited to acceptable measurements of trunk muscle strength and endurance (r = 0.23–0.66; Knapi, 1989). Sit-ups may involve varying accessory muscles besides abdominal muscles, such as the hip flexors. Therefore curl-up testing was selected to minimise the use of the hip flexors. While the reliability of dynamic or isometric sit-up tests seems to be limited, the curl-up test reached a good reliability (r = 0.92; Sparling et al., 1997). However, curl-up tests were criticized because it can be difficult to judge whether they are carried out correctly. The trunk muscle strength test (figure 1) could be an interesting alternative to measure global muscular strength and endurance of the trunk. It is a part of the standardised dynamical trunk muscle strength battery of the Swiss Olympic Medical Centres (Bourban et al., 2001; Tschopp et al., 2001). Its reliability was determined among athletes only (r = 0.87) and the authors judged the trunk muscle strength test to be valid to acquire health-related minimum requirements for elite athletes (Tschopp et al., 2001).

Motor skills are evaluated by determining test-retest reliability as for these test items a gold standard for validation is missing. The 1-leg standing test (OLS) is a feasible test to assess balance as a motor ability. Its interrater reliability (kappa value = 0.90) and the test-retest reliability (r = 0.73) are good and this test is further more valid for predicting ankle sprains in college students (Trojan and McKeag, 2006; Tsigilis et al., 2002). One limitation in evaluat-
Subject

In the reliability part of the study, 79 men (20.3 ± 1.1 y, 76.8 ± 13.5 kg, 179.9 ± 7.1 cm) completed all performance tests (twice). 60 men (20.3 ± 1.1 y, 76.7 ± 15.0 kg, 179.5 ± 6.6 cm) completed all performance tests in the validity part of the study. The feasibility of SPFTB was investigated among 1704 male draftees (19.5 ± 1.0 y, 72.7 ± 11.8 kg, 177.9 ± 6.5 cm).

Finally, standard values were obtained by the data of 15 794 conscripts who had to pass their recruitment to the Swiss Army. The entire fitness-test battery was completed by 81.4% of the conscripts who had to pass their recruitment to the Swiss Army. The body, other than the standing foot, had contact with the floor or the standing foot lost contact with the floor or the eyes were opened or the hands were released. For those who did not lose balance for 1 min, maximal time of 60 s was registered. Time was measured for both legs and valued with an accuracy of 0.1 s.

Dynamic balance was tested by walking forward and backward on a beam (WB) (length: 2.6 m, width: 0.1 m, height: 0.38 m). Subjects had to walk as fast as possible forward to the end of the beam and backward over the middle of it. Subjects who lost balance and descended were immediately showed where to step back on the beam. Time was measured with an accuracy of 0.1 s.

Validation of the test battery

Cardiorespiratory endurance

To validate PER, a 12-MRT and a VO₂peak-test were conducted. Previously, all subjects had performed the 12-MRT at least twice. 18 subjects (20.4 ± 1.3 y, 73.3 ± 10.8 kg, 180.0 ± 7.6 cm) from the validity part of the study were randomly selected to measure VO₂peak on a treadmill. These additional assessments were conducted during the following 3 weeks after the validation of the test battery. The start velocity was chosen individually, related to the performance in PER, between 8.5 and 14.0 km/h. Treadmill velocity was increased 0.5 km/h every min. Subjects were asked to run until exhaustion. VO₂ was measured with Oxycon Pro (Jäger, Hoechberg, Germany). The maximal value was recorded relative to body weight as VO₂peak. A linear regression between the peak running velocity of PER and VO₂peak of the treadmill test was calculated to estimate the VO₂peak of all subjects.

Muscle power

The results of SSP were validated against maximal power performance during free-weight bench press. Subjects had to push a 15 kg and a 30 kg barbell. The barbell was lowered slowly to touch the chest, held there for 1 s and then pushed to full arm extension as fast as possible. Force plates (MLD2, SPSport, Innsbruck, Austria) were attached under the bench to calculate the maximal power output during bench press. The best of 3 trials was registered for each weight.

Maximal power relative to body weight was assessed during SJ and CMJ on a force plate MLD2 and related to the results of SLJ. The performance of PS was compared to the running speed during a straight 40-m sprint. Light barriers recorded the sprint time between 30 and 40 m as a value for calculating the running speed. The best of 3 trials was valued with an accuracy of 0.001 s.

Trunk muscle strength

TMST was conducted after the protocol of Swiss Olympic Medical Centres (Tschopp et al., 2001) with specially trained physiotherapists, using a height adjustable positioning rack and head restraint.

Feasibility Questionnaire

A questionnaire was used in the feasibility part of the study. 25 responsible sport experts in recruitment centres were asked to rate the given statement, „the performance test is easily practicable”, for each performance test individually. In a second part, more feedback including information on the duration of the complete test battery was collected through open questions.

Data analysis

For every discipline, the measured values were directly used for data analysis, except for balance. The total score for balance was...
calculated adding the time of left and right OLS and then subtracting twice the time for WB (t_l + t_r – 2t_{WB}).

All statistical analyses were done with the program SPSS 13.0 (SPSS, Chicago, Illinois, USA). Pearson's correlation coefficients were calculated to estimate the relation between datasets. The t-test was used to estimate the significance of differences in repeated measurements. Descriptive analyses were done on the questionnaire data.

**Results**

**Physical performance tests**

**Cardiorespiratory endurance**

PER was highly reproducible (table 1). Additionally, PER time correlated positively with the distance run in the 12-MRT and with the VO_2peak in the maximal treadmill exercise test (table 2).

VO_2peak may be estimated by the peak running velocity of the PER with the following regression: VO_2peak [ml*kg\(^{-1}\)*min\(^{-1}\)] = 2.309 * Velocity\(_{peak}\) [km/h] + 16,549; (p<.001).

**Muscle power**

SSP yielded a good reliability (table 1). In 15 subjects, the force plates were not able to register the maximal power output during the bench press since the weight was never accelerated fast enough. SSP had a positive correlation with the maximal bench press power (table 2).

SLJ was highly reproducible (table 1). SLJ correlated positively with relative maximal power during normalised jumps. PS was inversely correlated with SLJ (table 2). While the performance of SLJ did not differ between both trials, the performance in PS was better in the second trial (table 1).

**Trunk muscle strength**

TMST yielded a good reliability (table 1). The simplified TMST and the original one correlated positively (table 2).

**Balance**

The repeated combination of static and dynamic balance tests generated a moderate correlation coefficient (table 1). The results in both cycles were not different. The retest of OLS alone yielded
a moderate reliability (table 1). OLS results tended to be better in the repeated test (table 1).

Feasibility

Based on the results of the reliability- and validity part of this study, PER, SLJ, SSP, TMST and bipart-balance test (OLS and WB) were selected for SPFTB. 3 sport experts needed less than 90 min to conduct the fitness test battery with 30 subjects, including information and warm up. The tests were rated as easily practicable or fairly easily practicable from 68% for SSP to 100% for SLJ. As an exception, the feasibility of WB was rated to be poor. In the open questions, 10 sport experts (40%) described WB as potentially dangerous or not well standardized.

Standard values

The representative SPFTB standard values (n = 12 862) for young men are presented in table 3.

Discussion

Physical performance tests

Cardiorespiratory endurance

Peak treadmill running velocity, during a speed-incremented VO2peak test, is an effective predictor of endurance performance (Harling et al., 2003). According to Noakes et al. (1990), peak treadmill running velocity is the best laboratory-measured predictor of running performance. The present study shows that peak running velocity in a progressive endurance run also reaches good reproducibility and validity if assessed as a field test on a track. PER was rated as feasible by sport experts. Additionally, PER may be more appropriate for individuals with low fitness levels or no experience in self-pacing compared to 12-MRT. However, further studies are needed to proof this assumption. We conclude that PER is an appropriate assessment of endurance capacity, especially for larger heterogeneous population groups.

Muscle power

SSP generated a good correlation with maximal power during bench press. Our results show that SSP is feasible, reliable and valid for young men. However, further research is needed to investigate the relationship between SSP of varying loads and the bench press power.

SLJ and PS are both valid and reliable. PS seems to be sensitive to learning effects, as the performance was better in the second measurement. This could be due to a learning effect in terms of agility. The comparison of PS with SLJ shows a strong relationship (table 2). Baker and Nance (1999) and Cronin and Hansen (2005) also found an inverse correlation between power and speed by comparing the jump height in CMJ with the 30 m sprint time (r = –0.56, p < 0.05) and the relative leg power with the 40 m sprint time (r = –0.76, p < 0.05). For an inexpensive physical performance test it is therefore reasonable to conduct either SLJ or PS. SLJ has previously been widely used and validated (r = 0.76; Markovic et al., 2004). Hence the power of lower extremities and running speed can be assessed with the reproducible, valid and feasible SLJ.

Trunk muscle strength

TMST is, due to its good reliability, validity and feasibility, suitable for use among larger groups of young men. However, the good validity has to be interpreted carefully because no gold standard for trunk muscle strength is available. In this study, TMST was compared with the established but more expensive TMST of the Swiss Olympic Medical Centres (Tschopp et al., 2001). We recommend using a height adjustable positioning-rack for anthropometrically heterogeneous groups.

Balance

To assess balance, a static and dynamic balance test was combined. This bipart-balance test is reproducible but WB seems not to be

<table>
<thead>
<tr>
<th>percentile</th>
<th>BMI [kg/m²]</th>
<th>PER [min/s]</th>
<th>pdt VO2peak [ml<em>kg⁻¹</em>min⁻¹]</th>
<th>SSP [m]</th>
<th>SLJ [m]</th>
<th>TMST [s]</th>
<th>OLS (t₁+t₄) [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>18.61</td>
<td>06:21</td>
<td>39.94</td>
<td>5.30</td>
<td>1.93</td>
<td>41.00</td>
<td>30.00</td>
</tr>
<tr>
<td>10</td>
<td>19.32</td>
<td>07:58</td>
<td>42.12</td>
<td>5.55</td>
<td>2.03</td>
<td>54.00</td>
<td>32.50</td>
</tr>
<tr>
<td>15</td>
<td>19.84</td>
<td>09:10</td>
<td>43.74</td>
<td>5.70</td>
<td>2.10</td>
<td>63.00</td>
<td>35.00</td>
</tr>
<tr>
<td>20</td>
<td>20.28</td>
<td>09:45</td>
<td>44.53</td>
<td>5.83</td>
<td>2.15</td>
<td>71.00</td>
<td>37.00</td>
</tr>
<tr>
<td>25</td>
<td>20.66</td>
<td>10:27</td>
<td>45.47</td>
<td>5.95</td>
<td>2.20</td>
<td>79.00</td>
<td>38.30</td>
</tr>
<tr>
<td>30</td>
<td>20.98</td>
<td>11:13</td>
<td>46.51</td>
<td>6.05</td>
<td>2.23</td>
<td>87.00</td>
<td>40.00</td>
</tr>
<tr>
<td>35</td>
<td>21.31</td>
<td>11:30</td>
<td>46.86</td>
<td>6.15</td>
<td>2.26</td>
<td>96.00</td>
<td>41.00</td>
</tr>
<tr>
<td>40</td>
<td>21.67</td>
<td>12:15</td>
<td>47.90</td>
<td>6.25</td>
<td>2.30</td>
<td>105.00</td>
<td>42.40</td>
</tr>
<tr>
<td>45</td>
<td>22.01</td>
<td>12:35</td>
<td>48.35</td>
<td>6.35</td>
<td>2.32</td>
<td>108.00</td>
<td>43.80</td>
</tr>
<tr>
<td>50</td>
<td>22.34</td>
<td>13:07</td>
<td>49.07</td>
<td>6.45</td>
<td>2.35</td>
<td>113.00</td>
<td>45.00</td>
</tr>
<tr>
<td>55</td>
<td>22.71</td>
<td>13:20</td>
<td>49.36</td>
<td>6.52</td>
<td>2.39</td>
<td>119.00</td>
<td>46.30</td>
</tr>
<tr>
<td>60</td>
<td>23.06</td>
<td>13:56</td>
<td>50.17</td>
<td>6.60</td>
<td>2.40</td>
<td>125.00</td>
<td>47.70</td>
</tr>
<tr>
<td>65</td>
<td>23.46</td>
<td>14:15</td>
<td>50.60</td>
<td>6.70</td>
<td>2.43</td>
<td>132.00</td>
<td>49.00</td>
</tr>
<tr>
<td>70</td>
<td>23.94</td>
<td>14:44</td>
<td>51.25</td>
<td>6.80</td>
<td>2.45</td>
<td>140.00</td>
<td>50.00</td>
</tr>
<tr>
<td>75</td>
<td>24.49</td>
<td>15:10</td>
<td>51.83</td>
<td>6.90</td>
<td>2.49</td>
<td>151.00</td>
<td>52.00</td>
</tr>
<tr>
<td>80</td>
<td>25.14</td>
<td>15:31</td>
<td>52.31</td>
<td>7.00</td>
<td>2.52</td>
<td>163.00</td>
<td>54.00</td>
</tr>
<tr>
<td>85</td>
<td>25.93</td>
<td>16:16</td>
<td>53.32</td>
<td>7.17</td>
<td>2.56</td>
<td>182.00</td>
<td>56.60</td>
</tr>
<tr>
<td>90</td>
<td>27.12</td>
<td>17:00</td>
<td>54.31</td>
<td>7.38</td>
<td>2.60</td>
<td>202.00</td>
<td>61.00</td>
</tr>
<tr>
<td>95</td>
<td>29.37</td>
<td>17:46</td>
<td>55.34</td>
<td>7.70</td>
<td>2.68</td>
<td>240.00</td>
<td>70.00</td>
</tr>
</tbody>
</table>

Table 3: Fitness-test battery standard values for young men (n = 12 862; 19.9 ± 1.0 y, 178.3 ± 15.9 cm, 72.8 ±12.0 kg). BMI, body mass index; SLJ, standing long jump; SSP, seated 2-kg-shot put; TMST, trunk muscle strength test; OLS, 1-leg standing; PER, progressive endurance run; pdt VO2peak, predicted peak oxygen consumption.
feasible. Therefore, we suggest using OLS alone. Our OLS is reproducible and a similar 1-leg balance test was shown to be valid for predicting ankle sprains (Trojan and McKeag, 2006). Therefore, only OLS was included in SPFTB. Further research is needed to find a feasible and reliable dynamic balance test.

**Limitations and strength of the study**

The subject’s motivation in the reliability- and validity part of this study may have been heterogeneous. It can be expected that results would be even better with highly motivated subjects. Especially, for the cardiorespiratory endurance tests, where 11 (reliability-part) and 16 (validity-part) subjects refused maximal performance, motivation is crucial.

2 different samples of subjects were used for the validity and reliability part of the study. This limitation was accepted in order to obtain more subjects for participation in the study. The 2 study groups are comparable. They do not differ in age, weight, height and all performance tests except SSP and SLJ (data not shown).

The balance test could not be validated because no gold standard to assess motor skills is available according to the authors’ knowledge. For a balance test a good reproducibility is already a challenge.

All measurements were done with a high number of subjects but exclusively with young men. Therefore, no data for other population groups are available so far. The replication of the present reliability, validity and feasibility studies for younger boys and girls will be the object of further research.

A major strength of the present study is that 3 important aspects of the fitness tests were assessed: reliability, validity and feasibility. In addition, standard values for future reference were collected with a representative sample among 19 year old men.

**Conclusion**

The new health- and performance-related SPFTB is qualified for nation wide use. SPFTB meets the previously specified demands; a) 3 sport experts are able to assess 30 subjects in 90 min, b) no expensive material is needed, c) the tests are valid and reliable and d) the most relevant components of fitness are included. With the 5 disciplines of SPFTB, changes in the physical fitness of specific population groups can be monitored. The new fitness-test battery may be attractive for epidemiological research, physical education, sport clubs and could be of interest in the field of occupational medicine (selection and control of employees in physically demanding jobs such as public service personnel and military).

Address for correspondence:

Thomas Wyss, Swiss Federal Institute of Sports, 2532 Magglingen, Tel. +41 (0)32 327 64 16, e-mail: thomas.wyss@baspo.admin.ch

**Literature**


